

University of Groningen

Influence of lung nodule margin on volume- and diameter-based reader variability in CT lung cancer screening

Han, Daiwei; Heuvelmans, Marjolein; Vliegenthart, Rozemarijn; Rook, Mienieke; Dorrius, Monique D; de Jonge, Gonda; Walter, Joan E; van Ooijen, Peter M A; de Koning, Harry J; Oudkerk, Matthijs

Published in:
British journal of radiology

DOI:
[10.1259/bjr.20170405](https://doi.org/10.1259/bjr.20170405)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Han, D., Heuvelmans, M., Vliegenthart, R., Rook, M., Dorrius, M. D., de Jonge, G., Walter, J. E., van Ooijen, P. M. A., de Koning, H. J., & Oudkerk, M. (2017). Influence of lung nodule margin on volume- and diameter-based reader variability in CT lung cancer screening. *British journal of radiology*, 90, [20170405]. <https://doi.org/10.1259/bjr.20170405>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Received:
31 May 2017

Revised:
21 September 2017

Accepted:
26 September 2017

<https://doi.org/10.1259/bjr.20170405>

Cite this article as:

Han D, Heuvelmans MA, Vliegenthart R, Rook M, Dorrius MD, de Jonge GJ, et al. Influence of lung nodule margin on volume- and diameter-based reader variability in CT lung cancer screening. *Br J Radiol* 2017; **90**: 20170405.

SCREENING SPECIAL FEATURE FULL PAPER

Influence of lung nodule margin on volume- and diameter-based reader variability in CT lung cancer screening

¹DAIWEI HAN, BSc, ^{1,2}MARJOLEIN A HEUVELMANS, MD, PhD, ^{1,3}ROZEMARIJN Vliegenthart, MD, PhD, ¹MIENEKE ROOK, MD, PhD, ^{1,3}MONIQUE D DORRIUS, MD, PhD, ³GONDA J DE JONGE, MD, PhD, ¹JOAN E WALTER, BSc, ^{1,3}PETER M A VAN OOIJEN, PhD, ⁴HARRY J DE KONING, MD, PhD and ¹MATTHIJS OUDKERK, MD, PhD

¹University of Groningen, University Medical Center Groningen, Center for Medical Imaging—North East Netherlands, Groningen, Netherlands

²Department of Pulmonology, Medisch Spectrum Twente, Enschede, Netherlands

³Department of Radiology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

⁴Department of Public Health, Erasmus University Rotterdam, University Medical Center Rotterdam, Rotterdam, Netherlands

Address correspondence to: Dr Marjolein A Heuvelmans

E-mail: m.a.heuvelmans@umcg.nl

Objective: To evaluate the influence of nodule margin on inter- and intrareader variability in manual diameter measurements and semi-automatic volume measurements of solid nodules detected in low-dose CT lung cancer screening.

Methods: 25 nodules of each morphological category (smooth, lobulated, spiculated and irregular) were randomly selected from 93 participants of the Dutch-Belgian Randomized Lung Cancer Screening Trial (NELSON). Semi-automatic volume measurements were performed using Syngo LungCARE® software (Version Somaris/5 VB10A-W, Siemens, Forchheim, Germany). Three radiologists independently measured mean diameters manually. Impact of nodule margin on interreader variability was evaluated based on systematic error and 95% limits of agreement. Interreader variability was compared with the nodule growth cut-off as used in Lung CT Screening Reporting and Data System (LungRADS; +1.5-mm diameter) and the Dutch-Belgian Randomized Lung Cancer Screening Trial (acronym: NELSON) /British Thoracic Society (+25% volume).

Results: For manual diameter measurements, a significant systematic error (up to 1.2 mm) between readers was found in all morphological categories. For semi-

automatic volume measurements, no statistically significant systematic error was found. The interreader variability in mean diameter measurements exceeded the 1.5-mm cut-off for nodule growth for all morphological categories [smooth: ±1.9 mm (+27%), lobulated: ±2.0 mm (+33%), spiculated: ±3.5 mm (+133%), irregular: ±4.5 mm (+200%)]. The 25% vol growth cut-off was exceeded slightly for spiculated [28% (+12%)] and irregular [27% (+8%)] nodules.

Conclusion: Lung nodule sizing based on manual diameter measurement is affected by nodule margin. Interreader variability increases especially for nodules with spiculated and irregular margins, and causes substantial misclassification of nodule growth. This effect is almost neglectable for semi-automated volume measurements. Semi-automatic volume measurements are superior for both size and growth determination of pulmonary nodules.

Advances in knowledge: Nodule assessment based on manual diameter measurements is susceptible to nodule margin. This effect is almost neglectable for semi-automated volume measurements. The larger interreader variability for manual diameter measurement results in inaccurate lung nodule growth detection and size classification.

INTRODUCTION

Interest in lung cancer screening by low-dose CT (LDCT) is increasing, ever since the National Lung Screening Trial (NLST) showed that LDCT screening for individuals at high risk of lung cancer reduces lung cancer mortality by 15–20%, compared with chest radiography.¹

A major drawback faced by the NLST was the high prevalence of false-positive screen results, which was 28.7% in NLST screen participants.² This high false-positive rate may result in patient harm and increased health-care cost. Improvements in nodule management such as a raised size threshold for positive nodules and the use

of nodule growth to identify malignant nodules have been suggested and were implemented in recent guidelines for lung nodule management.^{3–5}

Current measurement techniques used to assess the size of a nodule in LDCT screening in the USA rely on measurements of the maximum diameter or two maximum orthogonal diameters of a nodule by using electronic calipers. Lung CT Screening Reporting and Data System (Lung-RADS), a classification system proposed by American College of Radiology, and other current guidelines^{4,5} use the mean of maximum axial diameter and maximum perpendicular diameter of a single axial section (mean diameter) to determine the size of a nodule. In addition, Lung-RADS has defined nodule growth as a fixed increase of ≥ 1.5 mm in mean diameter. Nodule growth raises suspicion of malignancy and influences clinical management.

A number of European lung cancer screening trials have taken a different approach in nodule size assessment. Instead of manual diameter measurements, software for semi-automated measurement of nodule volume was used.^{6,7} While phantom studies have shown that nodules are at times over- or underestimated by semi-automated volume measurement compared with their true volume,^{7–13} this method offers better precision and reproducibility compared with manual diameter measurements.^{8,14} This is highly relevant in clinical practice, since greater reproducibility would result in increased sensitivity for nodule growth detection. According to the British Thoracic Society guidelines, the growth cut-off for lung nodule volume measurement is 25%.³

It can be hypothesized that diameter measurements perform particularly poorly in case of nodules with a non-smooth margin, although these nodules have the highest probability of malignancy.^{15,16} The influence of nodule margin on the precision of mean manual diameter measurements has not been studied before, and only limited data are available for the comparison between interreader variation of manual diameter measurements and semi-automatic volume measurements. The purpose of this study was to evaluate the influence of nodule margin on inter- and intrareader variability in manual diameter measurements and semi-automatic volume measurements of intermediate-sized solid nodules detected in LDCT lung cancer screening.

METHODS AND MATERIALS

Population and nodule selection

Data of the Dutch-Belgian Randomized Lung Cancer Screening Trial (NELSON), trial registration number: ISRCTN63545820, were used. The NELSON trial was approved by Ethics Committees of all participating centres, and authorized by the Dutch Healthcare Committee. All participants gave written informed consent. The design and conduct of the NELSON trial have been reported previously.^{17,18}

We randomly selected 100 intermediate-sized (50–500 mm³), non-calcified solid nodules found at baseline in lung cancer screening participants from the University Medical Center Groningen (Groningen, Netherlands) based on nodule-ID,

pre-stratified by nodule margin category (smooth, lobulated, spiculated and irregular). The number of samples per margin category (25 nodules) was defined by the number of nodules in the smallest category, to create subgroups with equal sample size. We selected only intermediate-sized nodules, since these nodules have the highest uncertainty regarding nodule nature, and usually lead to an extra short-term follow-up LDCT. In these nodules, it is of great importance that measurements lead to accurate evaluation of growth. Only solid nodules were included since LungCARE software (Siemens, Forchheim, Germany) is unable to semi-automatically calculate the volume of subsolid nodules.

CT scanning protocol

CT scanning was performed using 16-multidetector CT scanners (Sensation-16, Siemens Medical Solutions, Forchheim, Germany). All scans were realized in approximately 12 s in spiral mode with 16×0.75 mm collimation and 15-mm table feed per rotation (pitch, 1.5), in a cranial-caudal direction in low-dose setting, without intravenous contrast. Depending on body weight (<50, 50–80 and >80 kg), kVp settings were 80–90, 120 and 140 kVp, respectively. To achieve a CTDI-vol of 0.8, 1.6 and 3.2 mGy, respectively, the mAs settings were adjusted accordingly depending on the system used. To minimize breathing artefacts, scans were performed at inspiration with breath holding, after appropriate instruction of the participants. The images were reconstructed at a 1.0-mm slice thickness with a 0.7-mm increment. A medium-smooth B30f kernel was used for both detection and measurement of nodules.

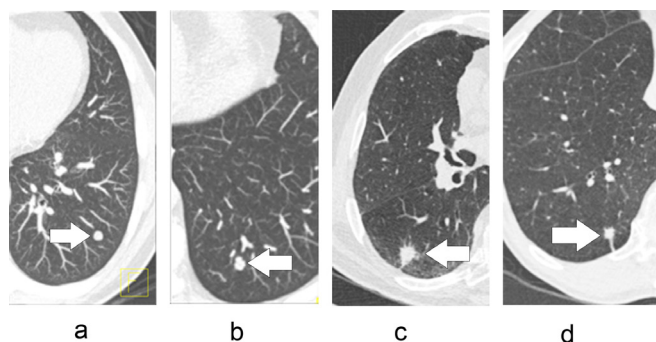
Image reading and measurements

Semi-automated volume measurements were performed using Siemens workstations with the Syngo LungCARE software package (Version Somaris/5 VB10A-W). Two sets of volumetric measurements done by independent radiologists from the NELSON trial were used in this study. Furthermore, for the purpose of this particular substudy, two chest radiologists and one abdominal radiologist with 8 (MDD), 7 (GdJ) and 6 (MR) years of experience in reading thoracic chest CT independently performed two sets of manual diameter measurements with at least 3 days between the two measurements, according to the Lung-RADS criteria (mean of the longest diameter and the longest perpendicular diameter). Rounding of diameter measurements was performed after calculation of the mean diameter according to Lung-RADS, as suggested by Li *et al*¹⁹ To perform the diameter measurements, images were uploaded to AquariusNET (Intuition Edition, v. 4.4.7, TeraRecon Inc, Foster City, CA) in random order, and images were read in lung window setting. The maximum axial diameter and maximum perpendicular diameter were measured using the caliper function. Mean diameter was defined as the mean of maximum axial diameter and maximum perpendicular diameter.

Nodule features

Nodules were defined as solid if their lung attenuation completely obscured the underlying structures.¹⁴ Based on the three-dimensional nodule segmentation derived from LungCARE, nodule margin was visually classified as smooth, lobulated,

Figure 1. Transversal CT image of four types of pulmonary nodules: (a) nodule with smooth margin, (b) nodule with lobulated margin, (c) nodule with spiculated margin and (d) nodule of irregular margin category.



spiculated or irregular, by the two independent NELSON radiologists who originally performed the volume measurements.¹⁸ In this classification, a smooth nodule had a smooth surface, a lobulated nodule had at least one abrupt bulging of the contour, a spiculated nodule had thicker strands extending from the nodule margin into the lung parenchyma without reaching the pleural surface, and an irregular nodule did not fit in one of the previous categories (Figure 1).^{19–21}

Statistics

To comply with the Lung-RADS criteria, rounded values of the mean diameter were used for nodule classification. For Bland–Altman analyses, non-rounded values were used. The inter- and intrareader agreement of nodule size measurements was examined for nodule subgroups (smooth, lobulated, spiculated and irregular) using the Bland–Altman method. An adapted Bland–Altman method proposed by Jones et al²² was used for the assessment of interreader agreement between three readers.²² Results from the analyses were presented as mean of absolute difference for manual diameter measurements and mean of relative difference for semi-automated volume measurements, with 95% limits of agreement (LoA). Relative difference was calculated as $(a - b)/m \times 100\%$, where a and b were measurements from two different readers and m was the mean of a and b . Diameter-based volume was calculated using mean axial diameter by assuming a spherical nodule shape, according to

$$V = \frac{1}{6} \cdot \pi \cdot D^3$$

with V = volume and D = mean axial diameter.

Friedman's test was used for comparisons between multiple readers. Wilcoxon signed-rank test was used for two-paired comparisons. Relative differences were compared against zero using the one-sample Wilcoxon signed-rank test.

The systematic error and 95%-LoA for volume and manual diameter measurement were compared with growth cut-offs: +25% for volume measurement based on the NELSON protocol and British Thoracic Society guidelines, and +1.5 mm for diameter measurements, based on Lung-RADS.^{3,4,23} Agreement in nodule size classification based on Lung-RADS

was analysed using Krippendorff's α where $\alpha \geq 0.8$ was considered as good, $0.8 > \alpha \geq 0.67$ as moderate and $\alpha < 0.67$ as poor agreement.^{24,25}

Parametric values were expressed as mean and 95% confidence interval (95% CI), non-parametric variables as median and interquartile range (IQR). A p -value < 0.05 was considered statistically significant. Statistical tests were performed using SPSS v. 22 (SPSS, IBM, New York, NY).

RESULTS

Median participant age was 59 years (IQR, 54–64), and participants' median smoking history was 39 pack-years (IQR, 30–47). 61 participants (66%) were current smokers. Median nodule volume was 118 mm^3 (IQR, 70–196 mm^3) and 116 mm^3 (IQR, 71–212 mm^3), determined by Reader 1 and Reader 2, respectively. Median mean nodule diameter was 6.7 mm (IQR, 5.7–8.3 mm) for Reader 1, 7.3 mm (IQR, 6.3–9.3 mm) for Reader 2 and 6.6 mm (IQR, 6.6–8.2 mm) for Reader 3. 26 nodules (26%) were attached to neighbouring anatomical structures such as pulmonary vessels, pleura and fissures (Table 1).

Systematic error

We found no significant systematic error for semi-automated volume measurements in the nodule subgroups as displayed in Table 2. For diameter measurements, both absolute and relative systematic error were statistically significant for at least one out of three reader comparisons for each nodule subgroup (Tables 2 and 3). A similar pattern as in the interreader analysis was found for intrareader comparison. However, for smooth nodules, we found no significant systematic error for the three readers ($p = 0.056$, $p = 0.957$ and $p = 0.116$) (Tables 2 and 4).

Inter- and intrareader variability and influence on growth and size classification

For interreader variability of volume measurements, the overall 95%-LoA was $\pm 23.7\%$, 5% below the 25%-growth cut-off (Table 2). For smooth and lobulated nodules, the 95%-LoA were ± 21.4 and $\pm 18.1\%$ (14.4 and 27.6% below the growth cut-off), respectively. The 95%-LoA of spiculated ($\pm 28.2\%$) and irregular nodules ($\pm 27.0\%$) exceeded the growth cut-off slightly, by 12.8 and 8.0%, respectively.

For interreader variability of manual diameter measurements, the overall 95%-LoA was ± 3.2 mm, exceeding the 1.5-mm growth cut-off by 113% (Table 2, Figure 2). The 95%-LoA exceeded the growth cut-off for all morphologies, the most for spiculated (± 3.5 mm) and irregular nodules (± 4.5 mm) for which the growth cut-off was exceeded by 133 and 200%, respectively. This resulted in an average of 10 (40%) growth misclassifications per pair of readers for spiculated nodules, and 9 (36%) growth misclassifications per pair for irregular nodules, based on diameter measurements. Also, intrareader variability of both spiculated nodules (± 2.4 mm) and irregular nodules (± 2.9 mm) exceeded the growth cut-off by 60 and 93%, respectively. For diameter-based volume, the 95%-LoA exceeded the 25% vol growth cut-off for both inter- and intrareader comparisons for all nodule margins.

Agreement in nodule categorization, based on Lung-RADS guidelines, was evaluated for the three readers (Table 5). There was consensus on nodule categorization in 56 smooth nodules (75%), 55 lobulated nodules (73%) and 46 spiculated and irregular nodules (61%). Moderate interreader agreement was found for mean diameter measurements of smooth [$\alpha = 0.67$, 95% CI (0.51, 0.81)], lobulated [$\alpha = 0.71$, 95% CI (0.59, 0.81)] and irregular [$\alpha = 0.72$, 95% CI (0.62, 0.82)] nodules. Poor interreader agreement was found for spiculated nodules [$\alpha = 0.5$, 95% CI (0.32, 0.67)]. Using *post hoc* analysis, interreader agreement was further evaluated. For spiculated nodules the Krippendorff's α coefficient remained poor (0.37–0.56). Overall, the Krippendorff's α varied between poor and moderate (0.54–0.74) for other nodule subgroups.

DISCUSSION

For this study, 100 intermediate-sized solid lung nodules of the NELSON trial's baseline round were selected randomly and assessed independently by three radiologists. We found significant intra- and interreader variation for manual mean diameter

measurements. In particular in non-smooth nodules, intra- and interreader variation was high, resulting in a moderate-to-poor interreader agreement in nodule categorization based on Lung-RADS. For semi-automatic volume measurements, interreader variability was affected by non-smooth nodule margins as well, although to a lesser extent than manual mean diameter measurements. For spiculated nodules, the 95%-LoA of semi-automatic volume measurements and manual diameter measurements exceeded the growth cut-off by 12 and 133%, respectively. Since nodule size and growth are the key discriminants to distinguish malignant from benign nodules in current guidelines, the measurement method with the smallest reader variability is preferable for nodule management in CT lung cancer screening.

Diameter measurements are commonly used in lung cancer screening studies and clinical practice. In the NLST, maximum axial diameter was used for size determination of detected lung nodules.¹ The Lung-RADS v. 1.0 (2014) and the updated guideline from the Fleischner society (2017) both recommend the use of mean diameter, since the average of long and short axis

Table 1. Information of nodule attachment and location

Nodule margin	Attachment type	Number of nodules	Nodule location	Number of nodules
Smooth	Pleural	6/25 (24%)	RUL	4/25 (16%)
	Vessel	0	RML	1/25 (4%)
	Intraparenchymal	19/25 (76%)	RLL	5/25 (20%)
			LUL	4/25 (16%)
			LLL	11/25 (44%)
Lobulated	Pleural	8/25 (32%)	RUL	8/25 (32%)
	Vessel	0	RML	1/25 (4%)
	Intraparenchymal	17/25 (68%)	RLL	2/25 (8%)
			LUL	9/25 (36%)
			LLL	5/25 (20%)
Spiculated	Pleural	2/25 (8%)	RUL	8/25 (32%)
	Vessel	0	RML	3/25 (12%)
	Intraparenchymal	23/25 (92%)	RLL	6/25 (24%)
			LUL	7/25 (28%)
			LLL	1/25 (4%)
Irregular	Pleural	7/25 (28%)	RUL	3/25 (12%)
	Vessel	3/25 (12%)	RML	3/25 (12%)
	Intraparenchymal	15/25 (60%)	RLL	5/25 (20%)
			LUL	11/25 (44%)
			LLL	3/25 (12%)
Total	Pleural	23/100 (23%)	RUL	23/100 (23%)
	Vessel	3/100 (3%)	RML	8/100 (8%)
	Intraparenchymal	74/100 (74%)	RLL	18/100 (18%)
			LUL	31/100 (31%)
			LLL	20/100 (20%)

LLL, left lower lobe; LUL, left upper lobe; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe.

Table 2. Absolute and percentage bias in nodule sizing between readers

	Systematic error (95%-LoA)							
	Interreader nodule volume		Interreader nodule diameter			Intrareader nodule diameter		
Nodule margins	Abs. (mm ³)	%	Abs. (mm)	DBV (mm ³)	%	Abs. (mm)	DBV (mm ³)	%
Smooth	0.7 (±22.4)	-0.1 (±21.4)	-0.1 ^a (±1.9)	-6.2 (±146.2)	-0.7 (±83.3)	0.0 ^a (±1.4)	2.8 (±90.0)	-0.1 (±71.0)
Lobulated	1.0 (±23.8)	-0.3 (±18.1)	-0.2 ^a (±2.0)	-29.6 ^a (±191.7)	-6.6 (±86.0)	0.2 ^a (±1.5)	21.6 ^a (±150.5)	-6.6 ^a (±66.4)
Spiculated	4.8 (±49.9)	-1.4 (±28.2)	0.0 ^a (±3.5)	-28.1 (±410.4)	3.9 (±122.1)	0.6 ^a (±2.4)	64.4 ^a (±278.0)	3.9 ^a (±83.9)
Irregular	0.7 (±61.3)	0.2 (±27.0)	0.5 ^a (±4.5)	77.5 ^a (±1100.0)	15.9 (±122.5)	0.4 ^a (±2.9)	89.7 ^a (±705.8)	15.9 (±84.9)
Total (n = 100)	-1.1 (±42.3)	-0.4 (±23.7)	0.0 ^a (±3.2)	3.4 (±602.3)	3.1 (±106)	0.3 ^a (±2.2)	44.6 ^a (±393.2)	10.5 (±78.0)

95%-LoA, 95% limits of agreement; Abs, absolute systematic error; DBV, diameter based volume.

% symbol denotes relative systematic error for volume and diameter-based volume.

Number in brackets is 95%-LoA, bold 95%-LoA exceeds the growth cut-off of 25% or 1.5 mm.

^aSignificant difference $p < 0.05$.

more accurately reflects three-dimensional nodule volume than the use of maximum diameter alone.^{4,5} In our study, the range of interreader variation in mean diameter of smooth and lobulated nodules was ± 1.9 and ± 2.0 mm (Table 2), exceeding the 1.5 mm growth cut-off as used in Lung-RADS by 27 and 33%, respectively. For spiculated and irregular nodules, the range of interreader variation was ± 3.4 and ± 4.5 mm, exceeding the 1.5 mm growth cut-off by 133 and 200%, respectively. Furthermore, according to the Lung-RADS classification, the mean diameter should be rounded to the nearest integer and, therefore, setting the growth cut-off at 1.5 mm seems unfit.

In clinical practice, when a lung nodule is detected, the CT will be compared with a previous CT. In case the nodule was present at a previous CT examination, often the mean diameters of the nodules at both scans are measured in one session to increase the

confidence in growth assessment. Thereby, interreader variability is excluded, but intrareader variability becomes an important influencing factor. Although the intrareader variability of diameter measurements in our study was smaller than the interreader variability, the 95%-LoA on intrareader variability for spiculated and irregular nodules still exceeded the 1.5 mm growth cut-off based on Lung-RADS guidelines (Table 3).

Since lung nodules at baseline are classified into different lung cancer risk categories based on size, consistent nodule size measurement between readers is important. In Lung-RADS, the diameter range of probably benign nodules is set to be ≥ 6 to < 8 mm, which falls within the range of measurement variation found in our study for spiculated and lobulated nodules, for both inter- and intrareader assessment. As a result, suspicious nodules (≥ 8 mm) could be misclassified as probably benign

Table 3. Systematic error for intrareader measurements by nodule margin

Nodule margin	Reader pairs					
	Systematic error (95% -LoA)					
	1 vs 2		1 vs 3		2 vs 3	
	Abs. (mm)	Rel. (%)	Abs. (mm)	Rel. (%)	Abs. (mm)	Rel. (%)
Smooth	-0.7 ^a (±1.4)	-11 ^a (±21)	-0.1 (±1.5)	-1 (±26)	0.7 ^a (±1.8)	11 (±29)
Lobulated	-0.5 ^a (±1.5)	-8 ^a (±24)	-0.3 (±2.1)	-3 (±30)	0.2 (±2.1)	4 ^a (±32)
Spiculated	-1.2 ^a (±2.7)	-14 ^a (±31)	0.0 (±3.2)	2 (±42)	1.2 ^a (±2.8)	17 (±39)
Irregular	0.1 (±5.9)	0 (±57)	0.7 ^a (±3.4)	9 ^a (±35)	0.6 (±3.8)	8 ^a (±39)
Total	-0.6 ^a (±3.5)	8 ^a (±37)	0.1 (±2.7)	2 ^a (±34)	0.6 ^a (±2.8)	10 ^a (±36)

95%-LoA, 95% limits of agreement; Abs, absolute systematic error; Rel (%), relative difference in percentage.

Number in brackets is 95%-LoA.

^aSignificant difference $p < 0.05$ (Wilcoxon).

Table 4. Systematic error for interreader measurements by nodule margin

Nodule margin	Systematic error (95%-LoA)					
	Reader number					
	1		2		3	
	Abs. (mm)	Rel. (%)	Abs. (mm)	Rel. (%)	Abs. (mm)	Rel. (%)
Smooth	-0.3 (± 1.5)	-5 (± 27)	0.1 (± 1.4)	1 (± 22)	0.2 (± 1.3)	3 (± 26)
Lobulated	-0.1 (± 1.3)	0 (± 19)	0.2 (± 1.0)	2 (± 17)	0.6 ^a (± 1.9)	8 ^a (± 29)
Spiculated	0.5 ^a (± 1.9)	7 ^a (± 27)	1.1 ^a (± 3.0)	14 ^a (± 36)	0.2 (± 1.9)	2 (± 26)
Irregular	0.8 ^a (± 2.7)	8 ^a (± 26)	0.3 (± 3.7)	2 (± 42)	0.1 (± 2.0)	1 (± 18)
Total	0.2 ^a (± 2.1)	2 (± 27)	0.4 ^a (± 2.6)	5 ^a (± 32)	0.3 ^a (± 1.8)	4 ^a (± 25)

95%-LoA, 95% limits of agreement; Abs, absolute systematic error; Rel (%), relative difference in percentage.

Number in brackets is 95%-LoA.

^aSignificant difference $p < 0.05$ (Wilcoxon).

(<6 mm) and vice versa. Using *post hoc* analysis, interreader agreement was further evaluated. Previous studies have shown that nodules with non-smooth margins have a higher probability of malignancy than nodules with smooth margins.^{16,26-29} Consequently, for these non-smooth nodules, small measurement variation and thus high consensus in their classifications is of even greater importance than for smooth nodules. Therefore, the use of maximum and mean diameter for assessing the size of small pulmonary nodules should be discouraged in lung cancer screening.

Studies evaluating the use of mean diameter for the assessment of small pulmonary nodules are limited. A previous study by Revel et al³⁰ found a 95%-LoA of ± 1.73 mm among three readers based on maximum diameter measurements, which once more is larger than the growth cut-off used in Lung-RADS. Unfortunately, in that study, influence of nodule margin was not analysed. In our study, the overall 95%-LoA among three readers was ± 3.2 mm. The larger variability might be explained by oversampling of spiculated and irregular nodules in our study (50%), compared

Table 5. Interreader agreement amongst three readers in Lung-RADS classification, by nodule margin

Nodule margin	α	95% CI	Observed agreement	Size category	Observed matrix			
Category					2	3	4A	4B
Smooth	0.67	0.51, 0.81	56 (75%)	2	20	6.5	0.5	0
				3	6.5	28.0	2.5	0
				4A	0.5	2.5	8	0
				4B	0	0	0	0
Lobulated	0.71	0.59, 0.81	55 (73%)	2	4	7	0	0
				3	7	33	3	0
				4A	0	3	18	0
				4B	0	0	0	0
Spiculated	0.50	0.32, 0.67	46 (61%)	2	1	4.5	0.5	0
				3	4.5	13	9.5	0
				4A	0.5	9.5	32	0
				4B	0	0	0	0
Irregular	0.72	0.62, 0.82	46 (61%)	2	3	5.5	0.5	0
				3	5.5	8	5.5	0
				4A	0.5	5.5	30	3
				4B	0	0	3	5

Lung-RADS, Lung CT Screening Reporting and Data System.

95% CI, 95% confidence interval of α ; Category 2, < 6 mm; Category 3, ≥ 6 -< 8 mm; Category 4A, ≥ 8 -< 16 mm; Category 4B, ≥ 15 mm; α , Krippendorff's α coefficient.

with prevalence of these nodules in the whole screening (9% of intermediate-sized nodules).³⁰

In a phantom study, Petrick et al³¹ found that the relative standard deviation of maximum diameter measurements of spiculated and elliptical nodules (20.3 and 16.4%) was larger compared with spherical and lobulated nodules (5.7 and 5.3%), while for semi-automated volume measurement the spiculated and elliptical nodules (8.3 and 3.6%) had similar relative SD as spherical and lobulated nodules (7.5 and 9.7%). This supports our finding that diameter measurement is more sensitive to the asymmetrical shape of a nodule than semi-automated volume measurement.

This study had some limitations. We focused on intermediate-sized nodules (50–500 mm³), since these nodules had highest uncertainty of nodule nature. As all nodules were classified indeterminate based on this volume, the interreader agreement for nodule size categorization could not be evaluated for the semi-automated volume measurements. Secondly, according to the Fleischner Society 2017 guidelines, mean diameter should be measured from the greatest dimension from transverse, coronal or sagittal reconstructed images. In our study, we used transverse reconstructed image for manual measurement. However, since the mean diameter measurement method stated in the Fleischner Society 2017 guidelines is still measured from a single reconstructed image, our results should be applicable to the new guidelines. Thirdly, in this study, volume measurements were performed with one specific software package. It should be kept in mind that volume measurements performed with any other software packages may be subject to different variations as we know from previous studies. Zhao et al³² have found significant differences in measured volume between LungCARE and two other software packages. The consensus on volume doubling time categorization was only 44–47% when LungCARE was compared with two other software packages. Owing to these variations in nodule measurements, we would like to advise to

perform subsequent nodule volume measurements on follow-up CT using the same software package. De Hoop et al³³ compared six software packages, including LungCARE. They showed that the 95%-LoA ranged from 16 to 22% for all software packages, similar to the 95%-LoA of 24% in the total of measured nodules reported in this study (Table 2). The performance of the different software packages in measuring nodules with different margins has not been studied yet. Lastly, we studied a relatively small number of nodules per nodule margin category. This was owing to the fact that only a limited number of irregular nodules were detected at baseline, and the total sample size was defined by the subgroup with smallest number of nodules.

In conclusion, we demonstrated that nodule assessment based on manual diameter measurements is susceptible to nodule margin. Interreader variability increases especially for nodules with spiculated and irregular margins. This effect is much smaller for semi-automated volume measurements. The larger interreader variability for manual diameter measurement results in moderate to poor classification of nodules based on their size, while growth misclassification may occur up to one third of cases. Therefore, semi-automated volume measurements are preferred over manual diameter measurements for nodule size and growth determination in CT lung cancer screening.

FUNDING

The NELSON-trial was sponsored by: Netherlands Organisation for Health Research and Development (ZonMw); Dutch Cancer Society Koningin Wilhelmina Fonds; Stichting Centraal Fonds Reserves van Voormalig Vrijwillige Ziekenfondsverzekeringen; Siemens Germany; Rotterdam Oncologic Thoracic Steering committee; G.Ph.Verhagen Trust, Flemish League Against Cancer, Foundation Against Cancer and Erasmus Trust Fund. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

REFERENCES

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; **365**: 395–409. doi: <https://doi.org/10.1056/NEJMoa1102873>
2. Pinsky PF, Gierada DS, Nath PH, Kazerooni E, Amorosa J. National lung screening trial: variability in nodule detection rates in chest CT studies. *Radiology* 2013; **268**: 865–73. doi: <https://doi.org/10.1148/radiol.13121530>
3. Callister ME, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015; **70**(Suppl 2): ii1–ii54. doi: <https://doi.org/10.1136/thoraxjnl-2015-207168>
4. Lung CT Screening Reporting and Data System (Lung-RADSTM). American College of Radiology. 2016. Available from: <http://www.acr.org/Quality-Safety/Resources/LungRADS> [cited 10 August 2016].
5. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on ct images: from the fleischner society 2017. *Radiology* 2017; **284**: 228–43. doi: <https://doi.org/10.1148/radiol.2017161659>
6. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; **361**: 2221–9. doi: <https://doi.org/10.1056/NEJMoa0906085>
7. Pedersen JH, Ashraf H, Dirksen A, Bach K, Hansen H, Toennesen P, et al. The Danish randomized lung cancer CT screening trial—overall design and results of the prevalence round. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer* 2009; **4**: 608–14. doi: <https://doi.org/10.1097/JTO.0b013e3181a0d98f>
8. Xie X, Zhao Y, Snijder RA, van Ooijen PM, de Jong PA, Oudkerk M, et al. Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detector CT: an anthropomorphic phantom study. *Eur Radiol* 2013; **23**: 139–47. doi: <https://doi.org/10.1007/s00330-012-2570-7>

9. Zhang L, Yankelevitz DF, Henschke CI, Jirapatnakul AC, Reeves AP, Carter D. Zone of transition: a potential source of error in tumor volume estimation. *Radiology* 2010; **256**: 633–9. doi: <https://doi.org/10.1148/radiol.10090924>
10. Das M, Ley-Zaporozhan J, Gietema HA, Czech A, Mühlenbruch G, Mahnken AH, et al. Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners. *Eur Radiol* 2007; **17**: 1979–84. doi: <https://doi.org/10.1007/s00330-006-0562-1>
11. Marten K, Funke M, Engelke C. Flat panel detector-based volumetric CT: prototype evaluation with volumetry of small artificial nodules in a pulmonary phantom. *J Thorac Imaging* 2004; **19**: 156–63.
12. Tao P, Griess F, Lvov Y, Mineyev M, Zhao B, Levin D, et al. Characterization of small nodules by automatic segmentation of X-ray computed tomography images. *J Comput Assist Tomogr* 2004; **28**: 372–7.
13. Way TW, Chan HP, Goodsitt MM, Sahiner B, Hadjiiski LM, Zhou C, et al. Effect of CT scanning parameters on volumetric measurements of pulmonary nodules by 3D active contour segmentation: a phantom study. *Phys Med Biol* 2008; **53**: 1295–312. doi: <https://doi.org/10.1088/0031-9155/53/5/009>
14. Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, Henschke CI. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. *Radiology* 2000; **217**: 251–6. doi: <https://doi.org/10.1148/radiology.217.1.r00oc33251>
15. Xu DM, van Klaveren RJ, de Bock GH, Leusveld A, Zhao Y, Wang Y, et al. Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. *Eur J Radiol* 2008; **68**: 347–52. doi: <https://doi.org/10.1016/j.ejrad.2007.08.027>
16. Xu DM, van der Zaag-Loonen HJ, Oudkerk M, Wang Y, Vliegthart R, Scholten ET, et al. Smooth or attached solid indeterminate nodules detected at baseline CT screening in the NELSON study: cancer risk during 1 year of follow-up. *Radiology* 2009; **250**: 264–72. doi: <https://doi.org/10.1148/radiol.2493070847>
17. Xu DM, Gietema H, de Koning H, Vernhout R, Nackaerts K, Prokop M, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; **54**: 177–84. doi: <https://doi.org/10.1016/j.lungcan.2006.08.006>
18. van Iersel CA, de Koning HJ, Draisma G, Mali W, Scholten ET, Nackaerts K, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer J* 2007; **120**: 868–74.
19. Li K, Yip R, Avila R, Henschke C, Yankelevitz D. P1.03-052 the effect of rounding on rate of positive results on CT screening for Lung Cancer. *J Thorac Oncol* 2017; **12**: S575. doi: <https://doi.org/10.1016/j.jtho.2016.11.723>
20. Takashima S, Sone S, Li F, Maruyama Y, Hasegawa M, Matsushita T, et al. Small solitary pulmonary nodules (< or =1 cm) detected at population-based CT screening for lung cancer: reliable high-resolution CT features of benign lesions. *AJR Am J Roentgenol* 2003; **180**: 955–64.
21. Takashima S, Sone S, Li F, Maruyama Y, Hasegawa M, Kadoya M. Indeterminate solitary pulmonary nodules revealed at population-based CT screening of the lung: using first follow-up diagnostic CT to differentiate benign and malignant lesions. *AJR Am J Roentgenol* 2003; **180**: 1255–63. doi: <https://doi.org/10.2214/ajr.180.5.1801255>
22. Jones M, Dobson A, O'Brian S. A graphical method for assessing agreement with the mean between multiple observers using continuous measures. *Int J Epidemiol* 2011; **40**: 1308–13. doi: <https://doi.org/10.1093/ije/dyr109>
23. Gietema HA, Wang Y, Xu D, van Klaveren RJ, de Koning H, Scholten E, et al. Pulmonary nodules detected at lung cancer screening: interobserver variability of semiautomated volume measurements. *Radiology* 2006; **241**: 251–7. doi: <https://doi.org/10.1148/radiol.2411050860>
24. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Commun Methods Meas* 2007; **1**: 77–89.
25. Krippendorff K. Reliability in content analysis: some common misconceptions and recommendations. *Hum Commun Res* 2004; **41**: 11–33.
26. Siegelman SS, Khouri NF, Leo FP, Fishman EK, Braverman RM, Zerhouni EA. Solitary pulmonary nodules: CT assessment. *Radiology* 1986; **160**: 307–12. doi: <https://doi.org/10.1148/radiology.160.2.3726105>
27. Zerhouni EA, Stitik FP, Siegelman SS, Naidich DP, Sagel SS, Proto AV, et al. CT of the pulmonary nodule: a cooperative study. *Radiology* 1986; **160**: 319–27. doi: <https://doi.org/10.1148/radiology.160.2.3726107>
28. Zwirerich CV, Vedral S, Miller RR, Müller NL. Solitary pulmonary nodule: high-resolution CT and radiologic-pathologic correlation. *Radiology* 1991; **179**: 469–76. doi: <https://doi.org/10.1148/radiology.179.2.2014294>
29. Zhao YR, Heuvelmans MA, Dorrius MD, van Ooijen PM, Wang Y, de Bock GH, et al. Features of resolving and nonresolving indeterminate pulmonary nodules at follow-up CT: the NELSON study. *Radiology* 2014; **270**: 872–9. doi: <https://doi.org/10.1148/radiol.13130332>
30. Revel MP, Bissery A, Bienvenu M, Aycard L, Lefort C, Frija G. Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? *Radiology* 2004; **231**: 453–8. doi: <https://doi.org/10.1148/radiol.2312030167>
31. Petrick N, Kim HJ, Clunie D, Borradaile K, Ford R, Zeng R, et al. Comparison of 1D, 2D, and 3D nodule sizing methods by radiologists for spherical and complex nodules on thoracic CT phantom images. *Acad Radiol* 2014; **21**: 30–40. doi: <https://doi.org/10.1016/j.acra.2013.09.020>
32. Zhao YR, van Ooijen PMA, Dorrius MD, Heuvelmans M, de Bock GH, Vliegthart R, et al. Comparison of three software systems for semi-automatic volumetry of pulmonary nodules on baseline and follow-up CT examinations. *Acta Radiol Stockh Swed* 1987. 2014; **55**: 691–8.
33. de Hoop B, Gietema H, van Ginneken B, Zanen P, Groenewegen G, Prokop M. A comparison of six software packages for evaluation of solid lung nodules using semi-automated volumetry: what is the minimum increase in size to detect growth in repeated CT examinations. *Eur Radiol* 2009; **19**: 800–8. doi: <https://doi.org/10.1007/s00330-008-1229-x>